



Next Generation Cloud Based Ingest & Processing Framework (I&PF) for Environmental Data

2017 AMS Annual Meeting

Rich Baker
Chief Architect
Solers, Inc.
Email: richard.baker@solers.com
Phone: 240-790-3338

Josh Leaverton, Peter MacHarrie, Dan Beall
Solers, Inc.

Dr. Shay Strong
OmniEarth, Inc.

www.solers.com

Cloud Based Ingest & Processing Framework (I&PF)

Uses Readily Available Open Source Technologies and Commercial Amazon Cloud Services



AUTO-SCALING



CODEDEPLOY



DYNAMODB



EC2



ELASTICCACHE



ELASTICSEARCH



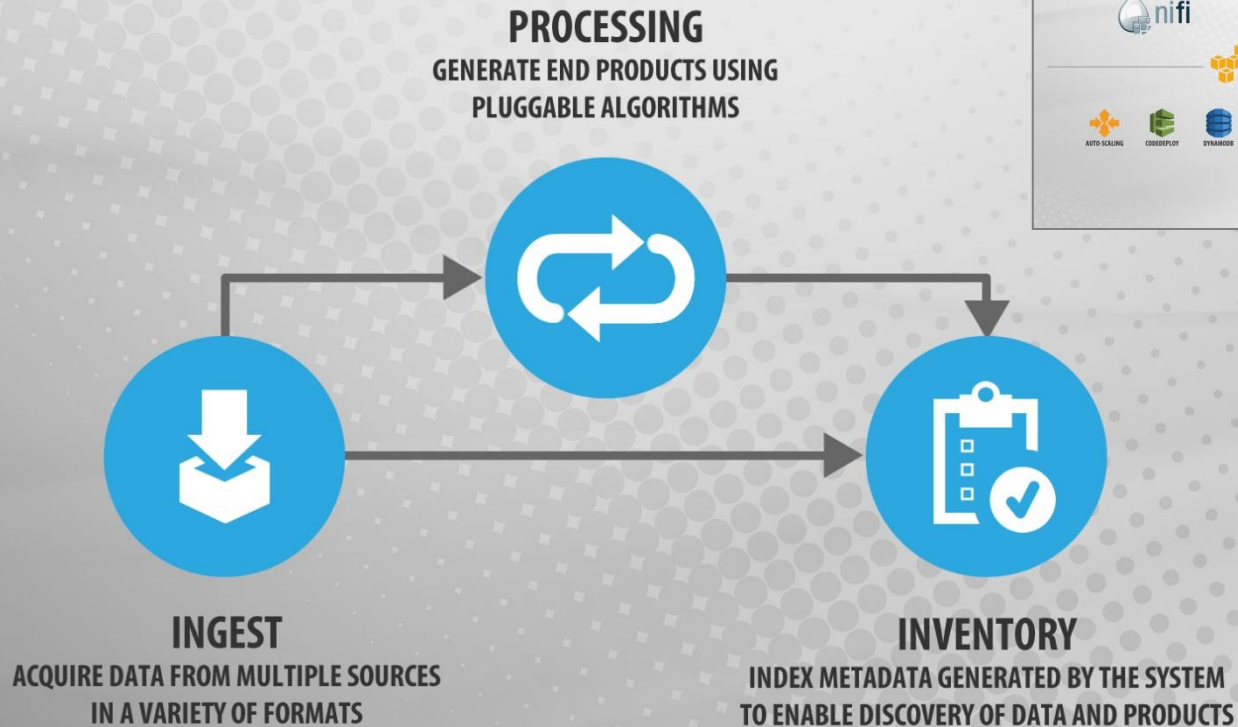
S3

OBJECTIVES:

- **Enable fast/easy integration** of data sources, product algorithms, and data consumers within a cloud based workflow (or “data pipeline”) framework
- Provide **easy to use web-based user interfaces** for discovery and access (for end users), as well as workflow monitoring and management (for algorithm developers and system operators/admins)
- Provide **RESTful web services** for other developers, scientists, etc. to **discover and access** the ingested/processed **data and metadata**, for use in other research / engineering initiatives (e.g., developing a new product algorithm)

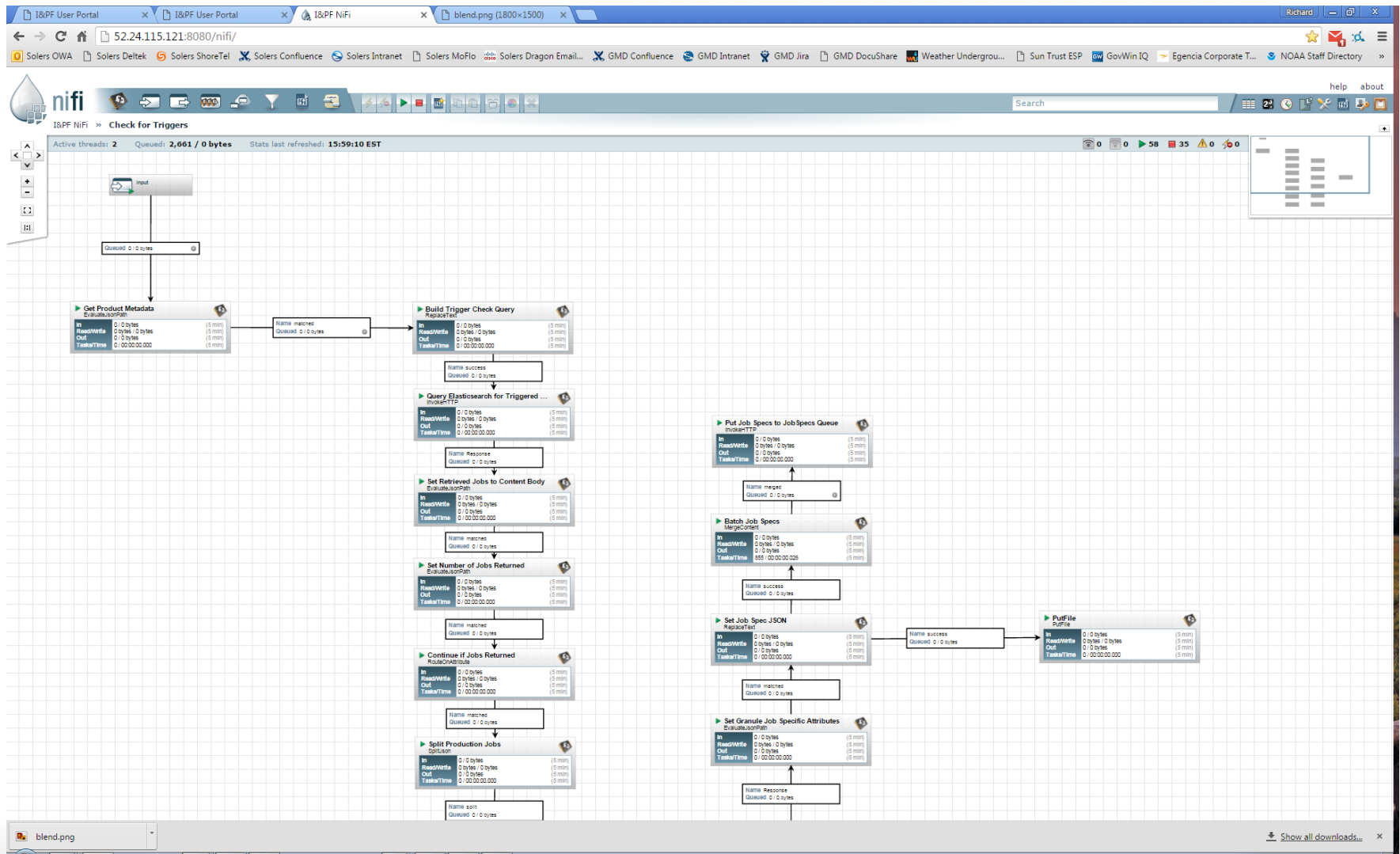
Cloud Based I&PF

High-Level Architecture



Cloud Based I&PF

Apache NiFi Workflow Engine



Cloud Based I&PF

3 NOAA Proof of Concept Use Cases (Solers IR&D)



➤ **NOAA S-NPP ATMS and MIRS**

- Ingests and inventories Suomi National Polar Partnership (S-NPP) Advanced Technology Microwave Sounder (ATMS) granules
- Generates Microwave Integrated Retrieval System (MIRS) products from the ATMS granules
- Makes ATMS granules and MIRS products searchable and accessible

➤ **NOAA Nexrad II Weather Radar**

- Ingests and inventories NOAA Nexrad II Weather Radar data sets that were published on Amazon S3 as part of the NOAA Big Data Project
- Makes NOAA Nexrad II Weather Radar data sets searchable and accessible

➤ **MIRS / Nexrad II Blended Product**

- Leverages the available MIRS products and NOAA Nexrad II Weather Radar data sets to produce a new blended product that combines the MIRS snow/water data with the Nexrad II radar data over mountainous regions

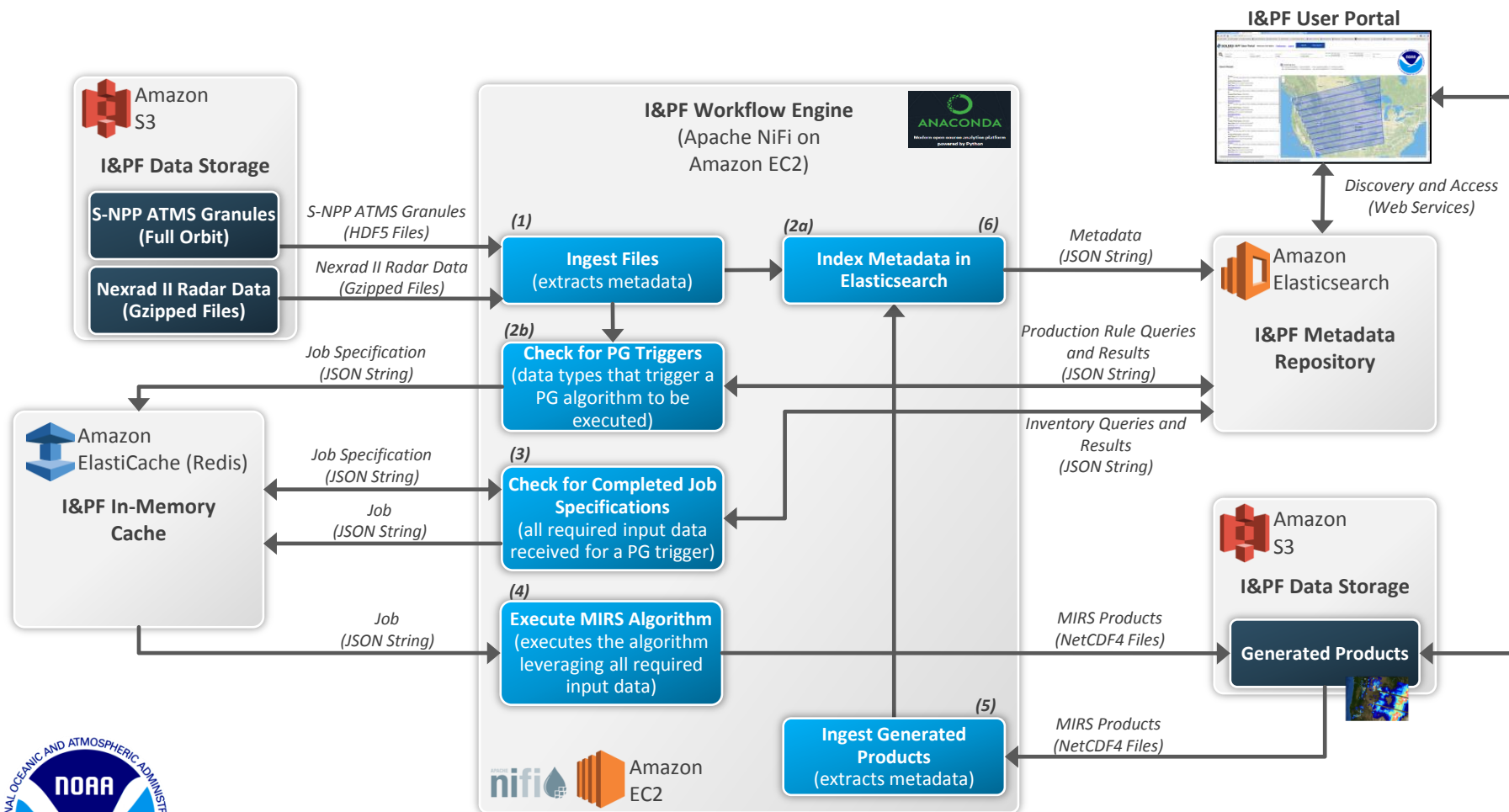
Cloud Based I&PF

NOAA Proof of Concept Architecture



Cloud-Based I&PF NOAA Proof of Concept

NOAA Data Ingest, Processing, and Inventory



NOAA Data/Products Made Available in the Cloud-Based I&PF: Ingested S-NPP ATMS Granules

I&PF User Portal | 52.24.115.121/ipf-portal/ | Richard

Solers OWA | Solers Detek | Solers ShoreTel | Solers Confluence | Solers Intranet | Solers MoFlo | Solers Dragon Email... | GMD Confluence | GMD Intranet | GMD Jira | GMD DocuShare | Weather Underground... | Sun Trust ESP | GovWin IQ | Egencia Corporate T... | NOAA Staff Directory

SOLERS I&PF User Portal | Welcome: Rich Baker | [Preferences](#) | [Logout](#) | [Search](#) | [Subscriptions](#)

Source Type: Satellite | Source: NOAA S-NPP | Instrument: ATMS | Product Short Name: ATMS-SDR | Start Date: mm/dd/yyyy | End Date: mm/dd/yyyy | Max Results: 10

Search Results

☒ Include Map Area

NW: 53.64463782485651, -128.6279296875 | NE: 53.64463782485651, -71.05957031249999
SW: 29.878755346037977, -128.6279296875 | SE: 29.878755346037977, -71.05957031249999

1. Product ID: SATMS_npp_d20151202_t1954330_e1955046_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:54:33.018Z
End Time: 2015-12-02T19:55:04.639Z
[Download Data Set](#)

2. Product ID: SATMS_npp_d20151202_t1954010_e1954326_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:54:01.018Z
End Time: 2015-12-02T19:54:32.639Z
[Download Data Set](#)

3. Product ID: SATMS_npp_d20151202_t1953290_e1954006_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:53:29.018Z
End Time: 2015-12-02T19:54:00.639Z
[Download Data Set](#)

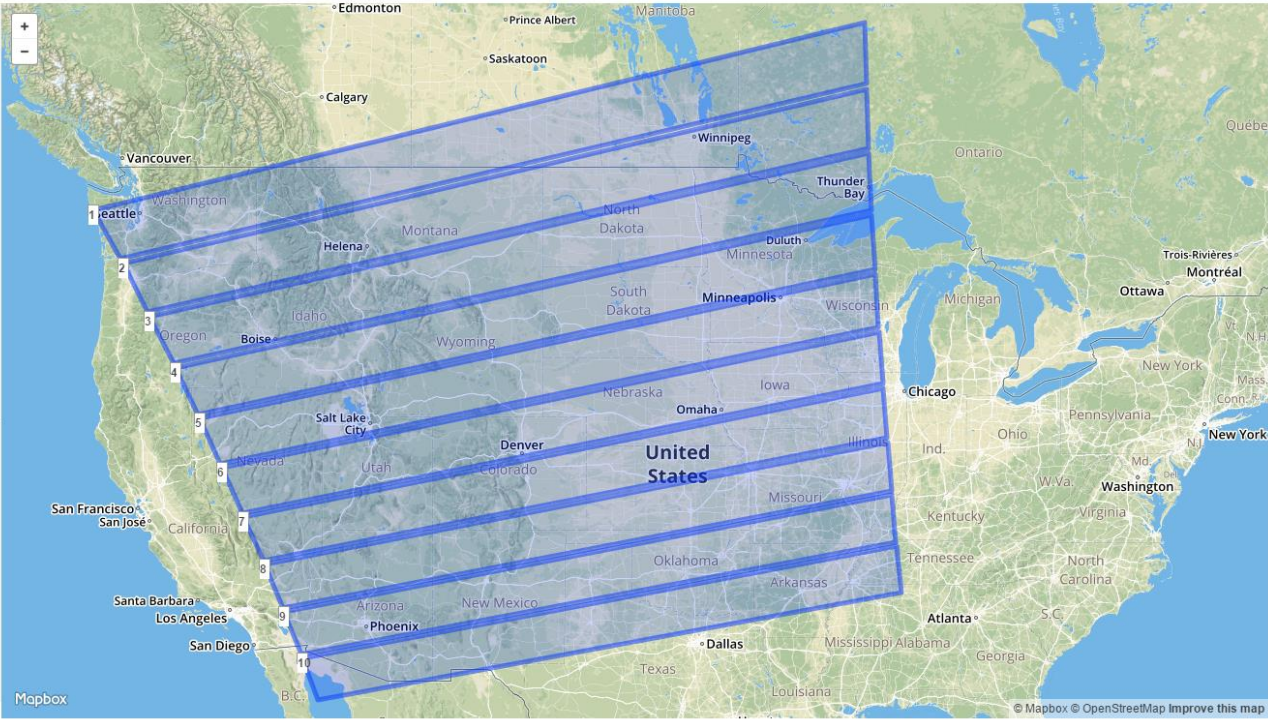
4. Product ID: SATMS_npp_d20151202_t1952570_e1953286_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:52:57.018Z
End Time: 2015-12-02T19:53:28.639Z
[Download Data Set](#)

5. Product ID: SATMS_npp_d20151202_t1952250_e1952566_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:52:25.018Z
End Time: 2015-12-02T19:52:56.639Z
[Download Data Set](#)

6. Product ID: SATMS_npp_d20151202_t1951530_e1952246_b21231_c20151214162
Product Short Name: ATMS-SDR
Start Time: 2015-12-02T19:51:53.018Z
End Time: 2015-12-02T19:52:24.639Z
[Download Data Set](#)

Mapbox

52.24.115.121/s3-ipf-data/products/ATMS-SDR/SATMS_npp_d20151202_t1954330_e1955046_b21231_c20151214162406256277_noaa_ops.h5



NOAA Data/Products Made Available in the Cloud-Based I&PF: Ingested Nexrad II Weather Radar Data

Browser tabs: I&PF User Portal, I&PF User Portal, I&PF NIFI, blend.png (1800x1500)

Browser address bar: 52.24.115.121/ipf-portal/

Browser bookmarks: Solers OWA, Solers Deltex, Solers ShoreTel, Solers Confluence, Solers Intranet, Solers MoFlo, Solers Dragon Email..., GMD Confluence, GMD Intranet, GMD Jira, GMD DocuShare, Weather Undergrou..., Sun Trust ESP, GovWin IQ, Egencia Corporate T..., NOAA Staff Directory

SOLERS I&PF User Portal Welcome: Rich Baker [Preferences](#) [Logout](#) [Search](#) [Subscriptions](#)

Search filters: Source Type: Radar, Source: NOAA Nexrad II, Instrument: Any, Product Short Name: Nexrad, Start Date: mm/dd/yyyy, End Date: mm/dd/yyyy, Max Results: 10

☒ Include Map Area

NW: 48.73445537176822, -128.47412109375 NE: 48.73445537176822, -99.68994140625
SW: 36.87962060502676, -128.47412109375 SE: 36.87962060502676, -99.68994140625

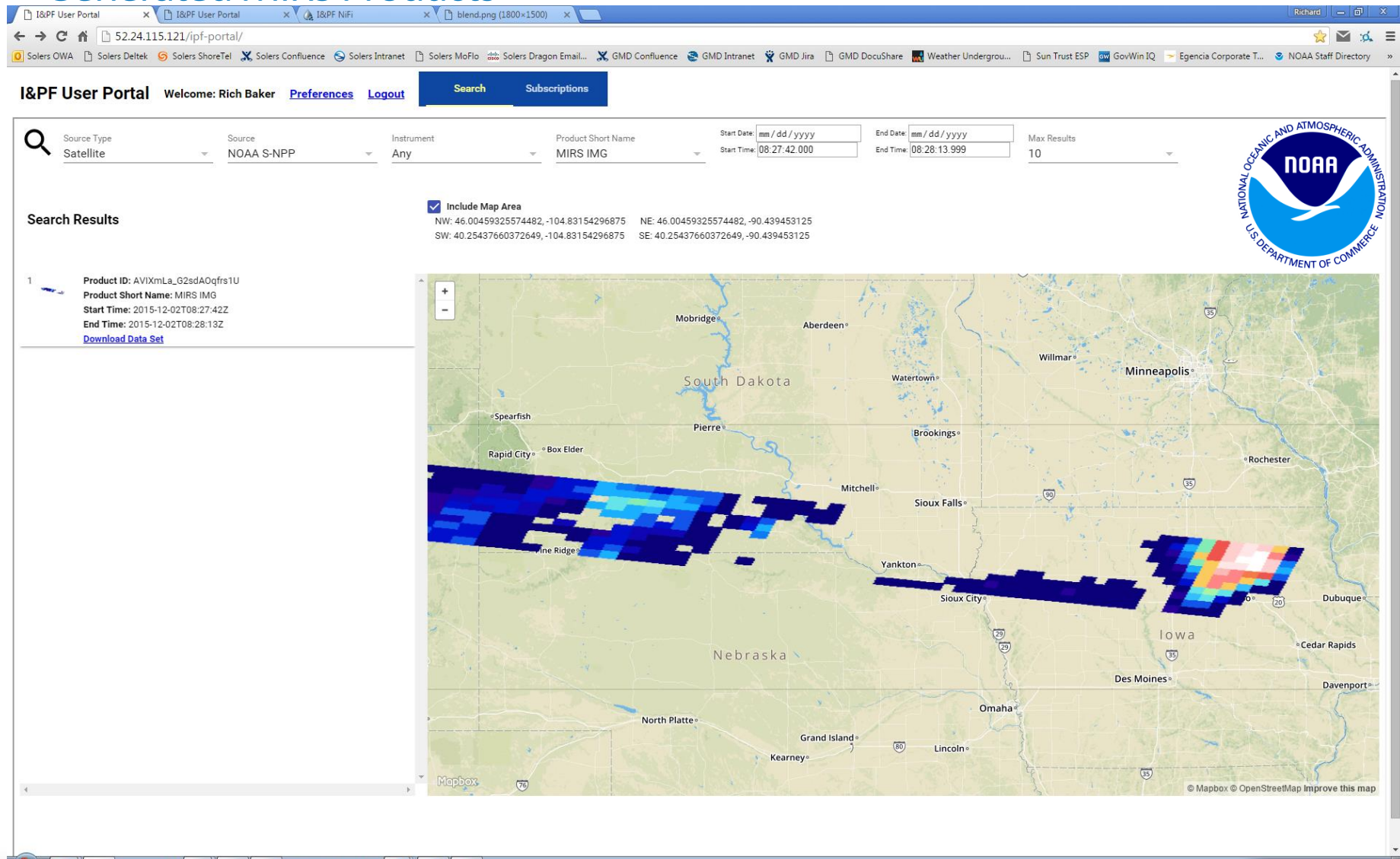
Search Results

Product ID	Product Short Name	Start Time	End Time	Download Data Set
1. KCBX20151202_235954_V06.gz	Nexrad	2015-12-02T23:59:54Z	2015-12-03T00:09:24.538Z	Download Data Set
2. KRTX20151202_235730_V06.gz	Nexrad	2015-12-02T23:57:30Z	2015-12-03T00:07:04.270Z	Download Data Set
3. KLRX20151202_235711_V06.gz	Nexrad	2015-12-02T23:57:11Z	2015-12-03T00:06:42.591Z	Download Data Set
4. KRGX20151202_235533_V06.gz	Nexrad	2015-12-02T23:55:33Z	2015-12-03T00:05:16.664Z	Download Data Set
5. KPDT20151202_235536_V06.gz	Nexrad	2015-12-02T23:55:36Z	2015-12-03T00:04:47.973Z	Download Data Set
6. KMAX20151202_235717_V06.gz	Nexrad	2015-12-02T23:57:17Z	2015-12-03T00:02:04.197Z	Download Data Set
7. KCBX20151202_235014_V06.gz	Nexrad	2015-12-02T23:50:14Z	2015-12-02T23:59:44.187Z	Download Data Set
8. KRTX20151202_234748_V06.gz	Nexrad			

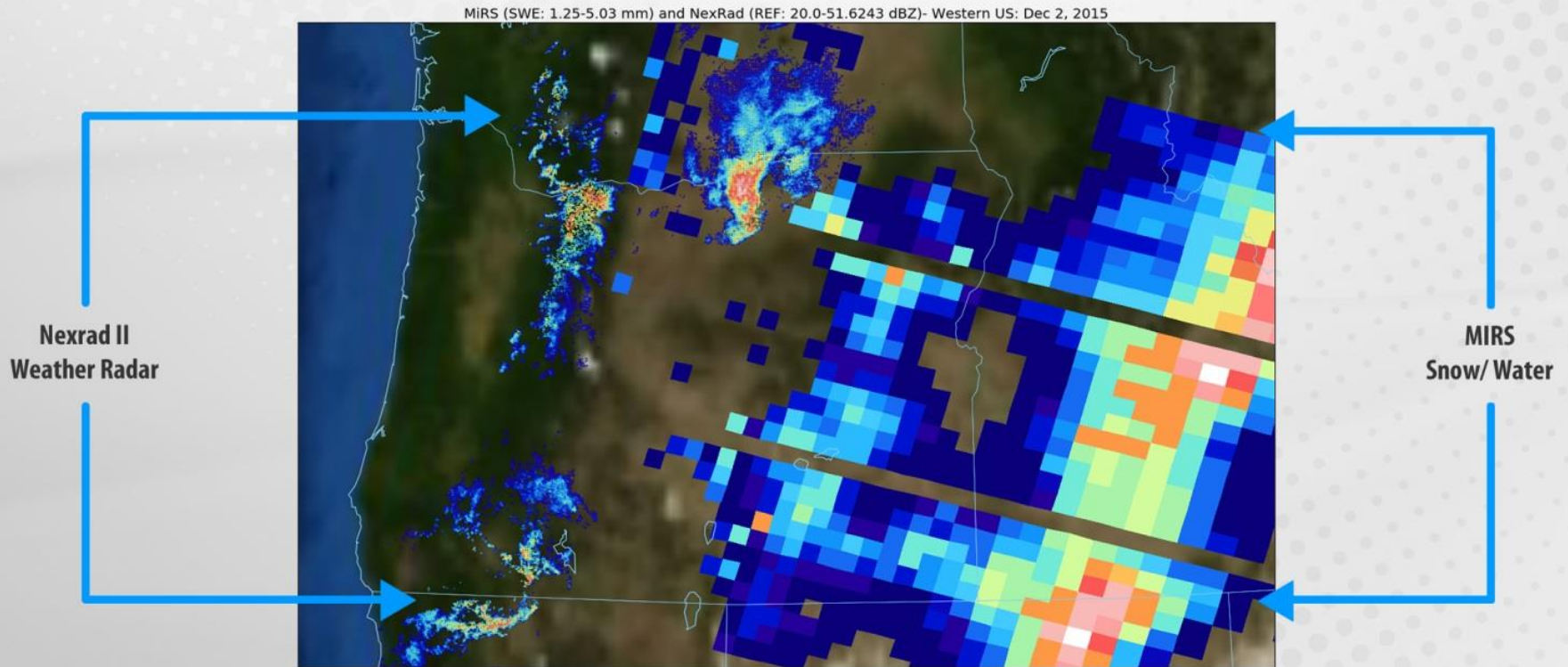
Map showing radar coverage areas (blue circles) over the Pacific Northwest and surrounding regions. The map includes labels for cities like Seattle, Portland, Salem, Medford, Eugene, San Francisco, San Jose, Reno, Carson City, Boise, Salt Lake City, Provo, Denver, and Cheyenne. The map is credited to Mapbox and OpenStreetMap.

NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce

NOAA Data/Products Made Available in the Cloud-Based I&PF: Generated MIRS Products



NOAA Data/Products Made Available in the Cloud-Based I&PF: MIRS / Nexrad II Blended Product



Cloud Based I&PF

OmniEarth Commercial Project

➤ OmniEarth Overview

- OmniEarth utilizes large satellite imagery sets combined with advanced machine learning algorithms to classify land cover for purposes of determining outdoor water budgets at the parcel level
- These budgets aid water agencies in drought-ridden communities in the US to best target water over-users

➤ Solers' OmniEarth Commercial Project

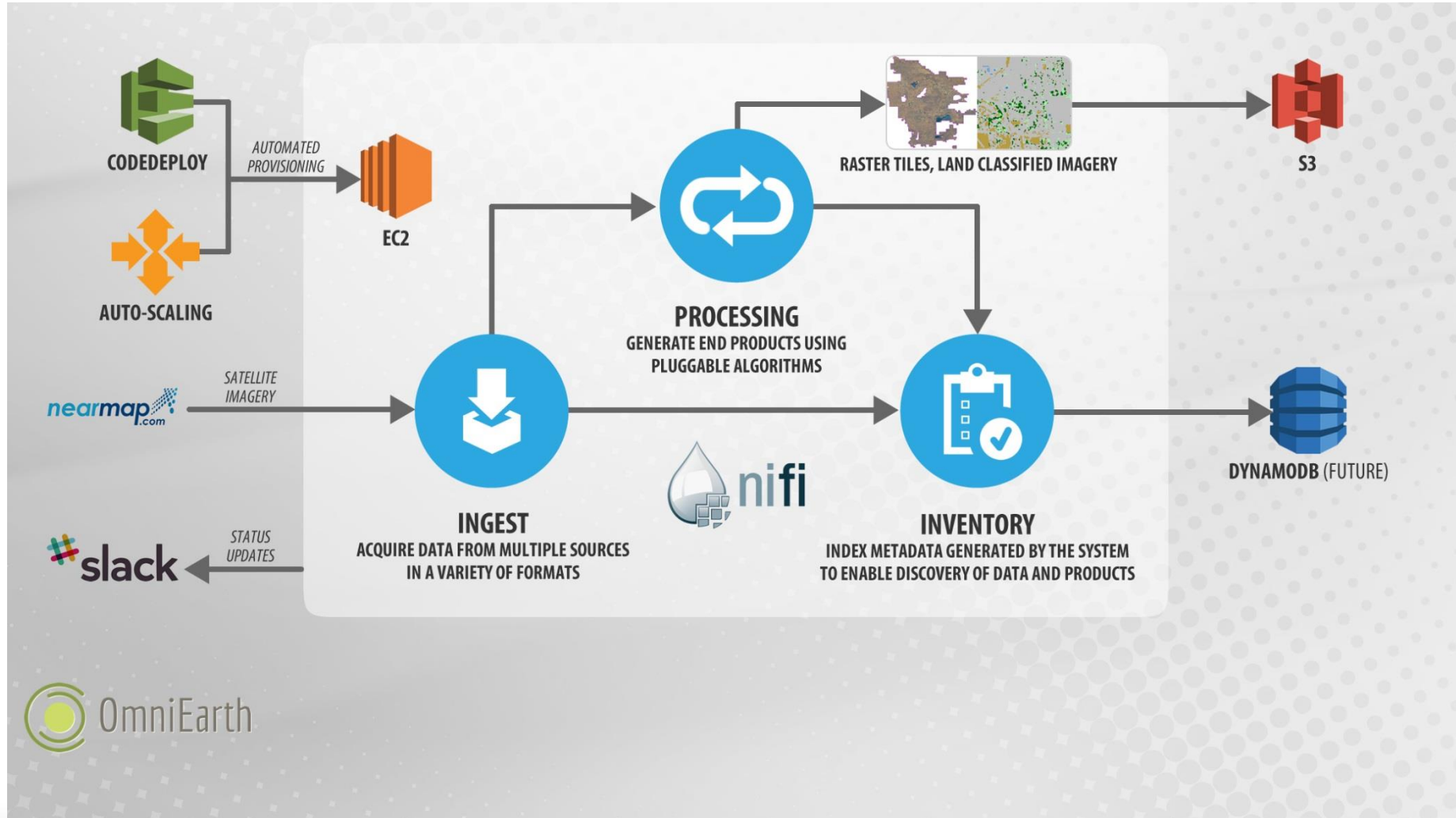
- Solers has partnered with OmniEarth's Data Scientists to help them utilize the Cloud Based I&PF in order to automate their (previously manual) satellite imagery ingest and land classification algorithm processing activities for their commercial Water Resource Management product



Water Resource Management Information:
<http://water.omniearth.net>

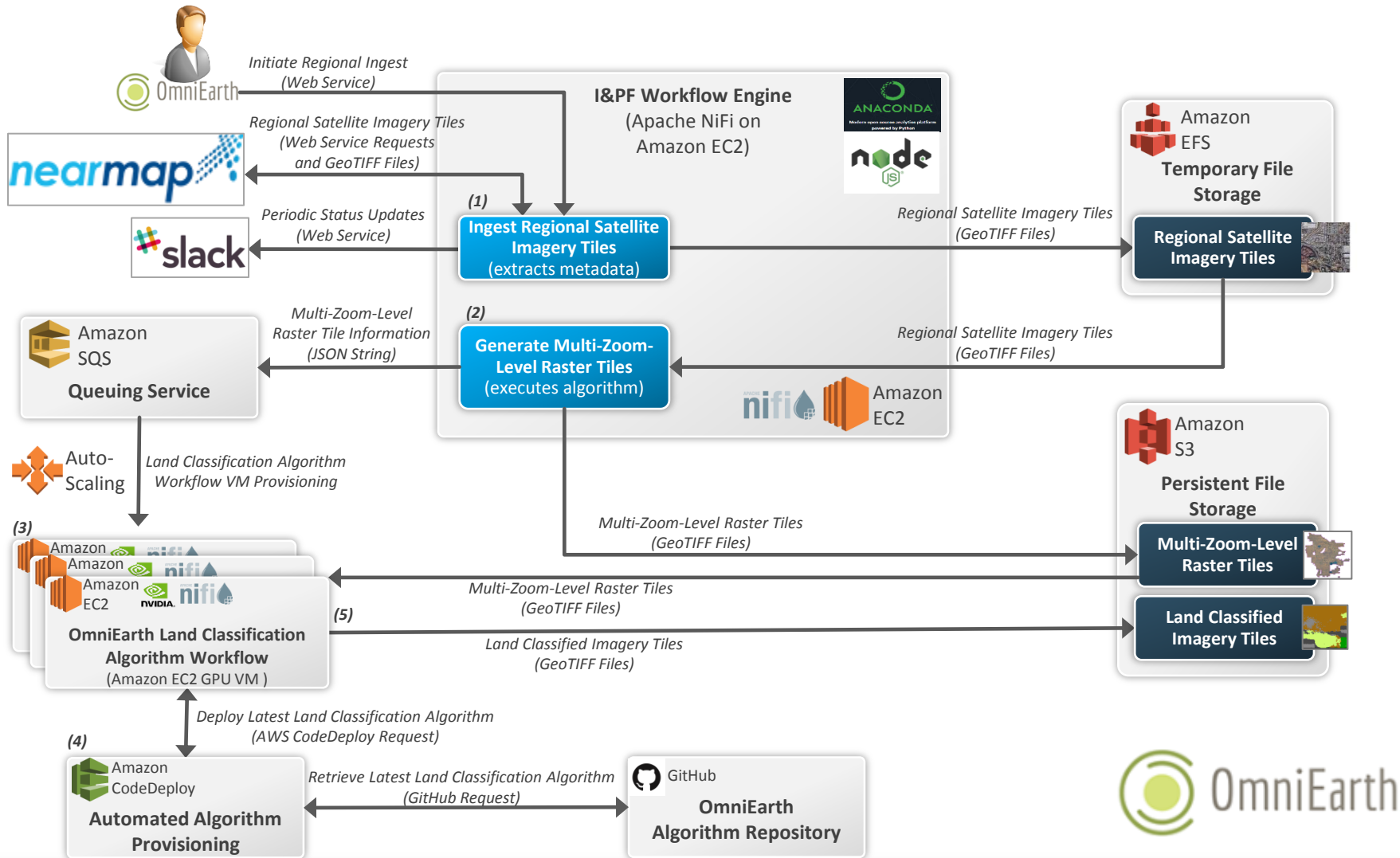
Cloud Based I&PF

OmniEarth Commercial Project Architecture



Cloud Based I&PF OmniEarth Commercial Project

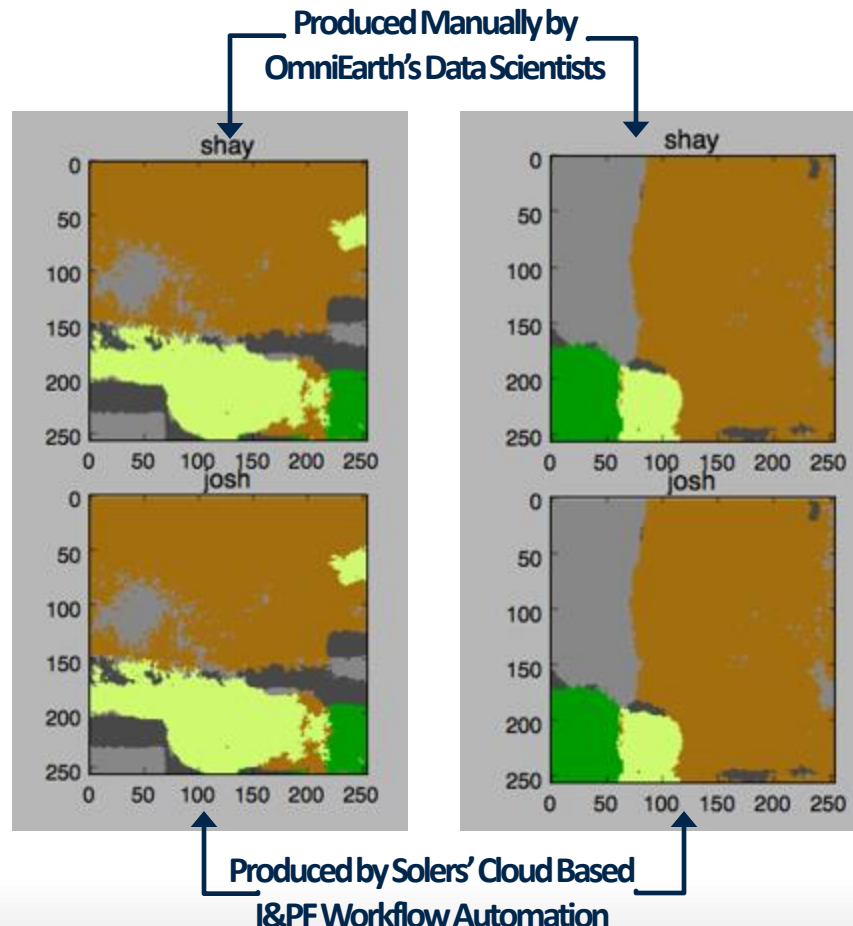
Satellite Imagery Ingest and Land Classification Processing



Cloud Based I&PF OmniEarth Commercial Project

Automation At Scale, Without Quality Reduction

Solers' Cloud-Based I&PF workflow automation produces land classified imagery tiles for OmniEarth, with the same level of precision and accuracy as those produced manually



Cloud Based I&PF OmniEarth Commercial Project

Outcomes and Benefits for OmniEarth

➤ Automation and Efficiency

- Automates their previously manual satellite data ingest and land classification processing activities
- Reduces the time to perform these activities by an order of magnitude (days to hours)
- Allows OmniEarth's Data Scientists to focus on improving their algorithms and training models, instead of manually running and watching over the satellite data ingest and land classification algorithm execution

➤ Tailored to the Data Scientist Needs

- Simple web service interface to initiate the workflow, based upon customer needs (e.g., specific regions of interest)
- Periodic monitoring/alerting of workflow status using a tool that is already heavily used and familiar to them (Slack)



Cloud Based I&PF

Future Utilities and Benefits

➤ **Development, integration, and test environment for Government (e.g., NOAA, NASA) satellite ground systems**

- Perform R&D and Cal/Val of new product algorithms for multiple satellites/platforms
- Scalable cloud-based framework that avoids on-premise infrastructure costs (pay just for the services that you need/use)
- Automation at scale with interfaces tailored to science algorithm developer and data scientist needs, helping to reduce the Research to Operations (R2O) timeline

➤ **Ingest and processing framework for commercial small satellite startup companies**

- Enable them to quickly get their satellite data ingested, processed, and available to users via a scalable cloud-based workflow or “data pipeline” framework, without requiring on-premise infrastructure

AMS 2017 Theme: Observations Lead The Way

➤ Your view on the greatest observational needs for your discipline in general

- Promotion of cloud-based platforms/frameworks for NOAA (FedRAMP approved, such as AWS GovCloud) to perform value-added capabilities with available observation data sets that are being published to cloud storage services (such as S3) as part of ongoing cloud-related initiatives (such as the NOAA Big Data Project)
- Index/catalog the data for discovery and access, and leverage cloud services to perform other value-added capabilities such as product generation, data assimilation, re-processing, etc. (more than just storing the data)

Questions

